

Situation Assessment and Hypothesis Testing in an Evolving Situation

M. A. Tolcott, F. F. Marvin, and T. A. Bresnick
Decision Science Consortium, Inc.

Research and Advanced Concepts Office
Michael Drillings, Acting Director

March 1996

19960604 069



United States Army
Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

Decision Science Consortium, Inc.

Technical review by

Michael Drillings

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE 1996, March	2. REPORT TYPE Final	3. DATES COVERED (from... to) September 1986-August 1989		
4. TITLE AND SUBTITLE Situation Assessment and Hypothesis Testing in an Evolving Situation		5a. CONTRACT OR GRANT NUMBER MDA903-86-C-0332		
		5b. PROGRAM ELEMENT NUMBER 0601102A		
6. AUTHOR(S) M. A. Tolcott, F. F. Marvin, and T. A. Bresnick (Decision Science Consortium, Inc.)		5c. PROJECT NUMBER B74F		
		5d. TASK NUMBER 1901		
		5e. WORK UNIT NUMBER C30		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Decision Science Consortium, Inc. 1895 Preston White Drive Reston, VA 22091		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-BR 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI		
		11. MONITOR REPORT NUMBER Research Note 96-34		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.				
13. SUPPLEMENTARY NOTES COR: Michael Drillings				
14. ABSTRACT (Maximum 200 words): This research investigated the effects of early judgment on (1) the handling of new information, some of which confirmed and some of which contradicted the early judgment, and (2) the selection of hypothesis-testing indicators. The context was situation assessment by Army intelligence analysts during an evolving battlefield scenario. Unaided analysts typically ignored or underweighted contradictory evidence; their confidence in their early judgment tended to rise. A second group was given a brief tutorial on common decision biases, and graphic displays that fostered awareness of uncertainty; in this group the tendencies were reduced (but not eliminated), and one-half of the group reversed their judgment at least once. A third group selected indicators; however, in the face of balanced feedback, their confidence remained constant rather than rising. The findings support the extension of confirmation bias theories to trained personnel performing realistic tasks. In addition, the results suggest that when decision makes the indicators they believe to be important, they pay more attention to contradictory evidence than when they are the passive recipients of new information.				
15. SUBJECT TERMS Decision Judgment Hypothesis testing Situation assessment Confirmation bias				
16. REPORT Unclassified			17. ABSTRACT Unclassified	18. THIS PAGE Unclassified
19. LIMITATION OF ABSTRACT Unlimited			20. NUMBER OF PAGES 57	21. RESPONSIBLE PERSON (Name and Telephone Number)
SECURITY CLASSIFICATION OF				

ACKNOWLEDGEMENTS

We gratefully acknowledge the support and encouragement of Dr. Michael Drillings, Basic Research Office, Army Research Institute.

The experiments would not have been possible without the cooperation of Dr. Julie Hopson, Chief, and Dr. Beverly Knapp, ARI Field Unit, Fort Huachuca, Arizona; and of LTC Okimoto and his predecessor, LTC Lamond, of the US Army Intelligence Center and School at Fort Huachuca. We appreciate their interest and cooperation. We are also especially grateful to the school's faculty, students, and staff who participated in the experiments.

EXECUTIVE SUMMARY

This research investigated the effects of early judgments on (1) the handling of new information, some of which confirmed and some of which contradicted the early judgments, and (2) the selection of hypothesis-testing indicators. The context was situation assessment by Army intelligence analysts during an evolving battlefield scenario. Unaided analysts typically ignored or underweighted contradictory evidence; their confidence in their early judgment tended to rise. A second group was given a brief tutorial on common decision biases, and graphic displays that fostered awareness of uncertainty; in this group the tendencies were reduced (but not eliminated), and half the group reversed their judgment at least once. A third group selected indicators they thought most important for testing their early judgments. They initially selected indicators that, if found, would tend to confirm their hypotheses, rather than selecting the most diagnostic indicators; however, in the face of balanced feedback, their confidence remained constant rather than rising.

The findings support the extension of confirmation bias theories to trained personnel performing realistic tasks. In addition, the results suggest that when decision makers select the indicators they believe to be important, they pay more attention to contradictory evidence than when they are the passive recipients of new information. Moreover, their subsequent hypothesis-testing strategies are more balanced. The practical implications of the findings for decision aiding, training, and operational procedures are discussed.

TABLE OF CONTENTS

	<u>Page</u>
1.0 INTRODUCTION	1
1.1 Objective	1
1.2 Background	1
2.0 PROCEDURES	4
2.1 General Description	4
2.2 Materials	5
2.3 Procedure	8
2.4 Participants	9
2.5 Evaluating the Collection Strategies	10
3.0 RESULTS	11
3.1 Initial Estimates and Confidence Levels	11
3.2 Selected Indicators	12
3.3 Overall Collection Strategies	14
3.4 The VOI for Various Collection Strategies	16
3.5 Effects of Feedback on Collection Strategy and VOI	20
3.6 Effect of Experience on VOI	22
3.7 Effect of Feedback on Confidence Level	24
4.0 CONCLUSIONS	29
4.1 Discussion of Findings	29
4.2 Recommendations	32
4.2.1 Training implications	32
4.2.2 Decision aiding implications	34
4.2.3 Research implications	37
REFERENCES	40
APPENDIX A	A-1

TABLE OF CONTENTS (cont'd)

LIST OF FIGURES

Figure	<u>Page</u>
3-1 Relationship Between General Collection Strategy and VOI	20
3-2 Percent of Maximum VOI Achieved as a Function of Years in Service	23

LIST OF TABLES

Table

2-1 Indicators and Their Likelihoods of Occurrence	7
3-1 Initial Estimates and Confidence Levels	11
3-2 Indicators Selected in Trial 1	13
3-3 Indicators Selected in Trial 2	13
3-4 Distribution of Indicators Selected	14
3-5 Trial 1 Strategies and VOI for Individual Participants .	17
3-6 Average VOI (Percent of Maximum) as a Function of Collection Strategy	20
3-7 Trial 2 Strategies and VOI for Individual Participants .	21
3-8 Trial 1 vs. Trial 2 Strategies and VOI	22
3-9 Relationship Between Experience and VOI Score	25
3-10 Allocation of Weights to Indicators	26
3-11 Effect of Feedback on Confidence Level	27
3-12 Trends in Average Confidence Levels, by Phase	27

1.0 INTRODUCTION

1.1 Objective

The general purpose of this research program has been to increase our theoretical understanding of human decision making in situations that evolve over time, and to derive implications for decision aids and training that might improve performance. More specifically, we have been concerned with the effects of early judgments on the handling of new information, some of which confirms and some of which contradicts the early judgment, and with the extent to which confidence in the early judgment is maintained or changed in the face of the new information. In addition, since much of the previous research on this problem has been conducted in academic settings with artificial problems, we were interested in determining the extent to which results of that research were found to hold with trained military personnel performing their occupational specialty task. In this research we have used Army military intelligence analysts assessing an evolving battlefield situation to estimate an enemy's most likely avenue of approach. The participants have been students, faculty and staff at the U.S. Army Intelligence Center and School, Fort Huachuca, Arizona. The exercise scenario and materials were adapted from an Army Central European battlefield scenario obtained from Fort Leavenworth, Kansas.

The program has been conducted in three phases, over a period of three years. This report presents the results of Phase 3, compares them with the Phase 1 and 2 findings, and derives recommendations based on the entire program.

1.2 Background

Phase 1 was essentially descriptive in nature; all participants, working in pairs, were given the same up-dated intelligence reports at intervals, and made their judgments without any special indoctrination or graphic decision aiding displays.

The results of Phase 1 (Tolcott, Marvin and Lehner, 1987; 1989) showed that regardless of the initial judgment, confidence was high and tended to increase

as the situation evolved. Confirming evidence was weighted significantly higher than disconfirming evidence. Only one pair of participants (out of 11) changed their initial hypothesis. Graphic rather than analytic approaches were typical, and base rates about enemy order of battle were largely ignored in resolving uncertainties. The Phase 1 findings are regarded as the baseline for interpreting the Phase 2 results.

In Phase 2, a comparable group of participants were given a brief description of commonly found decision biases and of the Phase 1 findings, and were provided with graphic displays to help them maintain an awareness of uncertainties as related to base rates and to foster their awareness of alternative hypotheses about enemy course of action. Results (Tolcott and Marvin, 1988) indicated a generally lower level of confidence, greater consideration of alternative hypotheses, and much more willingness to reverse early judgments based on new evidence. Half the teams (5 out of 10) changed their hypothesis at least once during the exercise. The tendency to overweight the importance of confirming evidence as compared with disconfirming, although not eliminated, was significantly reduced.

To summarize, the Phase 1 findings showed that trained personnel, working on problems in their area of expertise, can exhibit tendencies toward confirmation of early judgments and other non-normative cognitive behaviors similar to those found in laboratory tasks. In an evolving situation, their interpretation of the importance of new information can be influenced by models or schemata based on previous judgments. The findings of Phase 2 indicate that these tendencies can be reduced, although not entirely eliminated, by training innovations and by graphic aids that foster an awareness of uncertainty and provide help in dealing with it.

Phases 1 and 2 dealt with judgments by passive receivers of information; the term "confirmation bias" has been applied in this context to mean being relatively unresponsive to evidence against a favored hypothesis (e.g., Mynatt, Doherty, and Tweney, 1977; Einhorn, 1980; Einhorn and Hogarth, 1978), and we have used the term in that sense to describe our Phase 1 and 2 findings. "Confirmation bias" has been used in other senses as well. For example, Wason (1960; 1968) has used the term in the context of hypothesis

testing to mean choosing a question that is unlikely to falsify one's favored hypothesis. Baron, Beattie and Hershey (1988) further distinguish among several manifestations of confirmation bias, reserving the term "congruence bias" for the tendency to overvalue questions that have a high probability of a positive answer given the preferred hypothesis (see also Fischhoff and Beyth-Marom, 1983; Tweney, Doherty and Mynatt, 1982).

The active testing of hypotheses is a problem of significant importance in the intelligence community, where it is part of the more comprehensive activity known as collection management. Tendencies to seek evidence that would confirm a hypothesis could lead to serious gaps in information collected, and result in overlooking evidence highly diagnostic of enemy activity.

In view of the theoretical importance of hypothesis testing, and the practical importance of the confirmation bias in the intelligence context, it was decided to focus on active hypothesis testing in Phase 3 of this research program. The remainder of this report describes Phase 3, compare the results with the earlier findings, and presents recommendations based on the results of all three phases.

2.0 PROCEDURES

2.1 General Description

In all phases of this program, a Central European (Fulda Gap) battlefield scenario was presented to the participants, who were asked to play the role of intelligence analysts on the Divisional G-2 staff. In Phases 1 and 2, some participants were given a scenario in which the enemy deployment was weighted slightly to the north, some were given a scenario weighted south, and some were given a balanced scenario. They were read a summary of the events that followed the outbreak of hostilities three days previously. They were referred to maps and overlays during this reading, and were told that their task was to review these materials and the details that were available in the Intelligence and Order of Battle Workbooks, and estimate whether the enemy's main attack in the Divisional area would be in the northern or the southern sector, and to give their confidence level on a scale of 0-100%. In Phases 1 and 2, they were then given a set of new intelligence reports, some of which confirmed and some of which disconfirmed their early estimate, and some of which were neutral. They were then asked to give a new estimate and confidence level. This cycle was repeated two more times. At the end of the three cycles they were asked to review the intelligence reports and to rate each item in terms of the degree to which it confirmed or contradicted their initial hypothesis. The scenario and materials are described in detail in Tolcott, Marvin and Lehner (1987).

In Phase 3 the same basic maps, supplementary materials and instructions were used, except that all the participants received the same initial scenario, one in which the enemy deployment was balanced rather than weighted north or south. After giving their initial estimates of where the enemy main attack would occur, they were asked to select the indicators they would look for in order to check their hypothesis. They were allowed to select from a set of 16 indicators, and were told to pick the four that they thought would be the most valuable to them. They were then asked to rank the four in order of value, and to divide a total of 100 points among them, to indicate relative importance. They were then provided with feedback information on each of the four indicators; in two cases the feedback confirmed, and in two it

disconfirmed, their early estimate. They were then asked to give a new estimate and confidence level. This cycle was repeated once more, with participants selecting four of the remaining 12 indicators. The indicators from which they could make their selection were carefully prepared to represent differences in diagnosticity, frequency of occurrence, sector (north or south), and whether they were offensive or defensive indicators, as will be described in the next section.

The participants in Phase 3 were all officers (17 Captains, two First Lieutenants), all but one with military intelligence specialties, and all of whom were students in the Officer Advanced Course at the U.S. Army Intelligence Center and School (USAICS) at Fort Huachuca, Arizona. In Phase 3 the participants worked individually rather than in pairs; each session lasted about 1¹/₂ hours.

2.2 Materials

As mentioned earlier, the set of indicators from which the participants could make their selection was prepared in such a way that the indicators differed along several dimensions.

- 1) Diagnosticity: In general, a highly diagnostic indicator was one which had a high probability of occurrence if the enemy attack was in one sector, and a low probability of occurrence if the attack was in the other sector. Thus, "Forward movement of second echelon units in NORTH" was highly diagnostic, since its likelihoods were .95 if attack is north, and .30 if south, while "Decreased interception of radio traffic in NORTH" was not very diagnostic, with likelihoods that were .80 and .70, respectively.
- 2) Frequency of occurrence: Half the indicators were relatively common events and half were relatively unlikely.
- 3) Sector: For purposes of this experiment, it was assumed that any given information source could look only in one sector, north or south; thus, the indicators were sector-specific. For example, a participant could look for forward movement of heavy artillery units in the north, or in the south, or in both sectors; but in order to look in both sectors, two indicators would have to be selected.
- 4) Offensive vs. defensive: Half the indicators (e.g., movement of second echelon units) were offensive in nature, that is, they would be more likely to occur in the sector in which the attack was

occurring; others (e.g., preparation of minefields along FLOT, or preparation of field fortifications) were defensive, that is, they would be more likely to occur in the opposite sector.

The participant was given the probabilities for each indicator if the attack was in the North and if the attack was in the South, and was told that these probabilities were based on observations of enemy activities and knowledge of enemy doctrine. In general, the probabilities were considered reasonable. One exception was "Preparation of alternative artillery positions," which was assigned probabilities associated with a defensive indicator but which many participants nevertheless regarded as offensive.

Table 2-1 presents the indicators and probability (P) values as given to the participants in tabular form. (Note that the two columns at the right, which identify those indicators that are most diagnostic if seen and if not seen, were not given to the participants; they are presented here only for the reader's benefit.) They were also given the indicators in the form of a set of index cards, one indicator per card, as shown in the example at the bottom of Table 2-1.

The feedback information was given to the participants in the form of messages on cards. For each indicator, two feedback cards were prepared, one of which reported no evidence of the type being sought, while the other reported specific evidence of that type at a specific location. Thus, if a participant selected the indicator "Forward movement of heavy bridge units in NORTH," he would receive, as feedback, either a message saying "No forward movement of heavy bridge units in the northern sector," or a message saying "Convoy of six GSP ferries spotted moving west through Licherode (NB4152)," depending on whether he was to be given confirming or disconfirming feedback for that indicator (see Section 2.3, Procedure). Similarly, if he looked for movement of bridge units in the SOUTH, his feedback would be either "none" or "---spotted" at a specific location in the southern sector. It is important that the reader keep in mind the distinction between positive feedback (i.e., the sought-for evidence was seen) and confirming feedback (i.e., the feedback confirmed the initial hypothesis). Negative feedback can be confirming, and positive can be disconfirming.

Table 2-1: Indicators and Their Likelihoods of Occurrence

INDICATORS	IF ATTACK IS NORTH, LIKELIHOOD	IF ATTACK IS SOUTH, LIKELIHOOD	HIGHLY DIAGNOSTIC	
	IS	IS	IF SEEN	IF NOT SEEN
1 Preparation of minefields along FLOT in NORTH	0.05	0.60	X	
2 Preparation of field fortifications in SOUTH	0.95	0.60		X
3 Forward movement of heavy bridge units in NORTH	0.40	0.04	X	
4 Forward movement of heavy artillery units in SOUTH	0.40	0.80		
5 Forward movement of second echelon units in NORTH	0.95	0.30		
6 Forward movement of heavy artillery units in NORTH	0.80	0.40		
7 Forward movement of heavy bridge units in SOUTH	0.04	0.40	X	
8 Preparation of alternate arty postns in NORTH	0.15	0.40		
9 Forward movement of FROG missile units in SOUTH	0.10	0.50		
10 Forward movement of second echelon units in SOUTH	0.30	0.95		X
11 Preparation of minefields along FLOT in SOUTH	0.60	0.05	X	
12 Decreased interception of radio traffic in NORTH	0.70	0.80		
13 Forward movement of FROG missile units in NORTH	0.50	0.10		
14 Preparation of field fortifications in NORTH	0.60	0.95		X
15 Decreased interception of radio traffic in SOUTH	0.80	0.70		
16 Preparation of alternate arty postns in SOUTH	0.40	0.15		

Indicator:

Preparation of minefields along the FLOT in the northern sector.

If the main attack is in the southern sector, there is a 60% chance of this occurring.

If the main attack is in the northern sector, there is a 5% chance of this occurring.

2.3 Procedure

The sequence of events during each exercise session was as follows:

- 1) Investigator briefs participant on the general purpose the research (to study how people make decisions in an evolving situation), and assures him that this is not a test of his proficiency, that no names are used in our report, and that all data are statistically aggregated.
- 2) Investigator obtains background data on participant's MOS, schooling, years in service and in intelligence, and type of experience.
- 3) Investigator in role of G-3 Plans Officer at maps and overlays briefs participant on Corps situation and Commander's guidance, and identifies materials in the Intelligence and Order of Battle Workbooks.
- 4) Investigator points out first assumption, that errors in estimates are symmetrical; in other words that mistakenly predicting the main attack is in the north is just as bad for the Division as mistakenly predicting south. This assumption was necessary to avoid asymmetrical risk.
- 5) Participant completes analysis, and states initial estimate, confidence level, and reasons.
- 6) Investigator briefs participant on selection of intelligence indicators, and points out three more assumptions made for purposes of this experiment:
 - a. *Information sources are perfect.* That is, if you are told an event occurred, then it really did occur. If you look for certain evidence and are told it does not exist, then it really does not exist.
 - b. *Occurrence of events is not certain.* We know the probabilities of each event occurring under each of the two hypotheses. That is, for each piece of evidence there is some likelihood, not 100 percent, that it will occur if the enemy's main attack is in the north, and some likelihood that it will not occur. Also, there is a different likelihood that the same event would occur if the enemy's main attack is south. Thus, for each piece of evidence we will give you $P(E|N)$ (read as "probability of E given N") -- that is, the probability that the event would occur if the enemy's main attack is north; and $P(E|S)$ -- that is, the probability that it would occur if the main attack is south.
 - c. *Collection assets focus on areas.* Collection assets can look either in the northern or southern sector, but not across the

entire front. If you wish to look both north and south for the same event, you must select the two appropriate indicators.

- 7) Participant selects the four "most valuable" indicators from the set of 16 cards, arranges them in rank order, allocates 100 points among them to indicate relative value, and describes the reasons for his selection (i.e., his "collection strategy").
- 8) Investigator provides four cards showing results of collection effort. The feedback to each participant was balanced in the sense that feedback on two indicators was confirming (C) and two disconfirming (D). For nine participants, confirming evidence was given for the first and fourth indicators, and disconfirming for the second and third (i.e., the sequence was CDDC), while for ten participants the sequence was DCCD. This design was an attempt to equate as much as possible the potential importance associated with confirming and disconfirming evidence, the assumption being that the indicators ranked 1 and 4 would be approximately equal in importance to those ranked 2 and 3. The allocation of 100 points among the four indicators provided an independent check of this assumption.
- 9) Participant gives new estimate, confidence level, and reasons.
- 10) Participant selects four new indicators from the 12 remaining, arranges them in rank order, allocates 100 points among them, and states reasons.
- 11) Investigator provides feedback as in Step (8).
- 12) Participant gives final estimate, confidence level, and reasons.

2.4 Participants

Nineteen officers participated, of whom 17 were Captains and two First Lieutenants. The MOS of eight of them was 35D, a relatively new MI specialty in tactical intelligence; these eight were enrolled in courses in other MI specialties, such as signal intelligence. Ten of the participants had MOSs in other intelligence specialties (e.g., signal intelligence, counter-intelligence) and were enrolled in the 35D Officer Advanced Course. One participant was a field artillery specialist whose only experience in intelligence was in the 35D course. Years of active duty experience ranged from 3 to 16, with a mean of 7.2; years of experience in intelligence ranged from .25 to 12, with a mean of 6.2. Finally, 10 of the participants had had duty in Europe, and could be considered at least reasonably familiar with the battlefield environment in Germany.

2.5 Evaluating the Collection Strategies

To evaluate the collection strategies, we required a normative model with which to compare behavior. We used a Value of Information (VOI) model similar to the normative model described in Baron, Beattie and Hershey (1988). In this model, the value of a piece of information is defined as the extent to which it can reduce the expected error in the estimate. It is presumed that friendly troops would be deployed in accordance with the preferred hypothesis; therefore, the expected error is defined as the prior probability that the preferred hypothesis is not true. Two other assumptions have been made:

- (1) The "cost" (i.e., negative utility) of an error is the same in either direction;
- (2) The level of confidence expressed in the initial hypothesis is equivalent to the prior probability.

For example, if a participant expresses a 60% level of confidence in north as the sector of the enemy attack, 60% is taken as his prior probability of "north" being true, and 40% is the prior probability of it not being true (i.e., the expected error is 40%). Clearly, in this model the VOI depends (among other things) on the prior probabilities; the higher the prior for the preferred hypothesis, the less value any new evidence can have.

The other factors on which the VOI depends are: (1) the probability of the evidence given that the attack is in the north, $P(E|N)$; and (2) the probability of the evidence given that the attack is in the south, $P(E|S)$. Thus, the value of any item of information can be derived from the probabilities described above.

Since our procedure required that participants select items four at a time, the VOI had to be computed on that basis. The complete derivation of VOI for items taken four at a time is given in the Appendix.

3.0 RESULTS

3.1 Initial Estimates and Confidence Levels

As mentioned in Section 2.1, all the participants in the Phase 3 exercise were given exactly the same initial scenario, one in which the enemy forces were fairly equally balanced between the northern and southern sectors. Table 3-1 summarizes the findings regarding the initial estimates and confidence levels.

Table 3-1: Initial Estimates and Confidence Levels

<u>Initial Estimate</u>	<u>Number</u>	<u>Range of Confidence</u>	<u>Average Confidence</u>
North	14	55-100	76.1
South	5	70-80	76.0

As Table 3-1 indicates, 14 participants estimated "north" initially, while only 5 estimated "south." The most frequently given reasons for estimating an enemy attack in the north were that the West German units north of our sector were weak, that an enemy attack there would split the NATO forces along the boundary, and that the terrain (i.e., open spaces) provided a high-speed approach with good maneuverability for enemy tank forces. Most frequently given reasons for an attack in the south were that the enemy's probable strategic objective, Frankfurt, was to the south, that the enemy would exploit recent successes in the southern sector, and that the terrain (i.e., good road networks) provided a better high-speed approach for tank units. The difference of opinion as to what type of terrain is more favorable for a high-speed tank approach is worth noting; it suggests that analysis of terrain and enemy courses of action is highly personalized.

It is interesting to note that the average confidence levels were essentially the same (76.1 and 76.0) regardless of whether the initial estimate was "north" or "south." Remember that all participants received exactly the same initial scenario, one that was intended to be "balanced" in terms of enemy troop deployments and history of events during the previous three days. Since

the preponderance of initial estimates was "north," it would appear that the scenario slightly favored an attack in the north; however, the confidence level of those participants who estimated "south" was just as high as those who estimated "north." Although it is impossible to determine what an appropriate confidence level might be, these results strongly suggest that the participants were overconfident in their initial estimates. Indeed, one participant in the "north" group actually expressed a confidence of 100% in his estimate, the only such case observed in the course of the project. The findings suggest that more attention should be given during training to the concepts of uncertainty and probability.

3.2 Selected Indicators

It is of practical interest to examine which of the available indicators were actually chosen by the participants. Tables 3-2 and 3-3 show the frequency with which each indicator was selected in Trial 1 and Trial 2, respectively. In both tables a weighted total is also shown, arrived at by assigning 4 points if the indicator was selected first; 3 points is second; 2 points is third; and 1 point if fourth.

Table 3-2 shows that in Trial 1 there was a relatively heavy concentration on a few indicators, namely, movement of second echelon units and heavy artillery units in both sectors, and movement of bridge units in the north. Movement of second echelon units and of bridge units are both highly diagnostic indicators, according to the probabilities we provided and also according to enemy doctrine; however, heavy artillery movement would be a relatively likely occurrence in both the sector of attack and the other sector (.80 and .40 respectively), and therefore is not highly diagnostic. As shown in Table 3-3, the choices in Trial 2 were much more dispersed. In part this was due to the fact that, for most of the participants, their preferred indicators had been used in Trial 1 (they could not be used twice), but as will be shown later, it also reflects a change in collection strategy, namely, a tendency to begin looking for defensive indicators. The collection strategies will be discussed in more detail in the next section.

Table 3-2: Indicators Selected in Trial 1

<u>Indicator</u>	# of Times Selected				<u>Total</u>	Wtd <u>Total</u> *
	<u>1st</u>	<u>2nd</u>	<u>3rd</u>	<u>4th</u>		
1. Minefields North	1	1	1	0	3	9
2. Field fortifications South	0	0	0	2	2	2
3. Bridge units North	1	2	3	4	10	20
4. Heavy artillery units South	1	3	3	1	8	20
5. Second echelon units North	10	3	2	2	17	55
6. Heavy artillery units North	2	4	5	2	13	32
7. Bridge units South	0	0	1	1	2	3
8. Alternate arty pos'ns North	0	0	0	0	0	0
9. FROG missile units South	0	0	0	1	1	1
10. Second echelon units South	4	5	2	0	11	35
11. Minefields South	0	0	0	4	4	4
12. Decreased radio traffic North	0	0	1	1	2	3
13. FROG missile units North	0	0	0	0	0	0
14. Field fortifications North	0	1	1	0	2	5
15. Decreased radio traffic South	0	0	0	1	1	1
16. Alternate arty pos'ns South	0	0	0	0	0	0

Table 3-3: Indicators Selected in Trial 2

<u>Indicator</u>	# of Times Selected				<u>Total</u>	Wtd <u>Total</u> *
	<u>1st</u>	<u>2nd</u>	<u>3rd</u>	<u>4th</u>		
1. Minefields North	1	3	2	0	6	19
2. Field fortifications South	1	4	3	2	10	24
3. Bridge units North	1	2	1	2	6	14
4. Heavy artillery units South	0	3	1	0	4	11
5. Second echelon units North	1	0	1	0	2	6
6. Heavy artillery units North	2	0	0	0	2	8
7. Bridge units South	1	0	2	1	4	9
8. Alternate arty pos'ns North	1	2	0	6	9	16
9. FROG missile units South	0	1	1	3	5	8
10. Second echelon units South	3	0	2	0	5	16
11. Minefields South	3	1	3	0	7	21
12. Decreased radio traffic North	0	0	0	1	1	1
13. FROG missile units North	4	0	1	0	5	18
14. Field fortifications North	1	1	0	3	5	10
15. Decreased radio traffic South	0	1	1	1	3	6
16. Alternate arty pos'ns South	0	1	1	0	2	5

*Note: 4 points if selected 1st
 3 points if selected 2nd
 2 points if selected 3rd
 1 point if selected 4th

3.3 Overall Collection Strategies

Of primary interest in this phase of the research were the collection strategies used by the participants, and in particular, how their actual strategies compared with the theoretically optimum strategy as determined by the Value of Information model. The data were analyzed to examine the extent to which the participants sought (a) information in the sector in which they had hypothesized the enemy attack, as compared with the non-hypothesized sector, (b) indicators that were evidence of enemy offensive activity, as compared with defensive, and (c) indicators that would tend to confirm their hypothesis, as compared with those that would disconfirm it. Note these characteristics are not independent: an offensive indicator in the hypothesized section would be confirming, as would a defensive indicator in the non-hypothesized sector. Finally, it is of interest to compare the collection strategies used during the first trial (i.e., selection of the first four indicators) with those used in the second trial, after feedback had been received.

Table 3-4 presents three sets of frequencies showing distribution of indicators selected in Trials 1 and 2; (a) shows the indicators in the hypothesized (H) vs. non-hypothesized (\bar{H}) sectors, (b) the offensive (Off) vs. defensive (Def) indicators, and (c) the confirming (Con) vs. disconfirming (Dis) indicators. A Chi-square test showed that all the differences between column totals are significant, and that the differences between trials are also significant ($P = <.01$).

Table 3-4: Distribution of Indicators Selected

	(a)		(b)		(c)	
	H	\bar{H}	Off	Def	Con	Dis
Trial 1	54	22	63	13	57	19
Trial 2	30	46	30	46	37	39
Totals	84	68	93	59	94	58

The first thing to notice is that, as far as totals are concerned, there was a strong tendency to choose indicators in the hypothesized sector, that were offensive rather than defensive, and that if found would confirm rather than disconfirm the initial hypothesis. A closer examination of the tables shows that this tendency is entirely due to the strategies selected in Trial 1; the data for Trial 2 show a reversal of this effect, although the reversal is not strong enough to overcome the differences in the totals.

From a practical point of view, the data in Table 3-4(a) are perhaps of greatest concern, since they suggest that, if faced with a choice of areas to search, intelligence analysts would tend to focus on the area in which they believe the enemy will attack, ignoring evidence in other areas that might be even more diagnostic. It is true that for purposes of this exercise it was assumed that any single collection asset could search in only one sector, an assumption that may not be completely realistic. But the evidence is strong that, if resources were limited, they would probably be deployed mainly in the sector where the enemy attack is expected, a form of confirmation bias that could seriously degrade the quality of the situation assessment. It is noteworthy that in Trial 2, after receiving balanced feedback in Trial 1, the participants tended to look more in the non-hypothesized sector; this shift in strategy was overtly described in the verbal protocols as due to having received inconclusive evidence in Trial 1, and wanting to check on enemy activities in the non-hypothesized sector. Often indicators sought in the non-hypothesized sector in Trial 2 were defensive in nature, which if found would tend to confirm the hypothesis, but sometimes the search was for offensive indicators in the non-hypothesized sector, to avoid the possibility of missing some diagnostic evidence.

Table 3-4(b) shows that the indicators selected in Trial 1 were overwhelmingly offensive in nature, with a tendency to reverse this strategy in Trial 2. Several reasons were given for this: (a) enemy doctrine emphasizes offensive rather than defensive activities; (b) offensive activities would occur early and therefore should be looked for first, because of the lead time required to move offensive units into position; and (c) offensive indicators were perceived as more diagnostic than defensive ones, despite the probabilities given. The reasons for reversing the strategy in Trial 2 were that (a) time

had elapsed so that defensive activities were more likely to be taking place, and (b) the indicators perceived to be most diagnostic had already been selected in Trial 1. The collection strategy with regard to offensive and defensive indicators appears to be justified by at least some of these reasons.

The tendency in Trial 1 to look for offensive indicators in the hypothesized sector means that the preponderance of evidence sought in Trial 1 was confirming rather than disconfirming, as shown in Table 3-4(c). Furthermore, there is only a slight tendency to reverse this trend in Trial 2, where the number of confirming indicators selected (37) was almost as great as the number of disconfirming (39). Therefore, the disparity in the totals is greatest for this characteristic. This finding supports much of the research in this field, which has found a general tendency to seek evidence that would confirm rather than disconfirm a hypothesis.

In only four cases did a participant's verbal explanation of the strategy used indicate that the probabilities played a significant role. These four participants recognized the fact that the spread between the two probabilities was related to the diagnosticity of the indicator. In one of these cases it was acknowledged that a mixed strategy was used, that is, one that combined diagnosticity with a desire to obtain confirming evidence. In the other 15 cases, the verbal records suggest that when a highly diagnostic indicator was chosen, it was more because of knowledge of enemy doctrine than because of the probabilities themselves. This provides us with assurance that our probabilities were for the most part realistic.

3.4 The VOI for Various Collection Strategies

It has already been shown that there was a significant tendency, in Trial 1, to select indicators that were in the hypothesized sector, offensive rather than defensive in nature, and confirming rather than disconfirming of the initial hypothesis. These tendencies, if pursued in actual operations, might result in the overlooking of diagnostic indicators. The question remains as to how effective these strategies are when measured against the normative VOI model.

Table 3-5 presents the relevant Trial 1 data for each participant individually. The columns, from left to right, show:

confidence in the initial hypotheses (the prior probability);
 number of indicators chosen in the hypothesized sector;
 number of offensive indicators;
 number of confirming indicators;
 maximum VOI achievable for the assessed prior probability;
 actual VOI achieved;
 percent of maximum VOI achieved.

Table 3-5: Trial 1 Strategies and VOI for Individual Participants

<u>Prior Prob.</u>	<u># Hyp.</u>	<u># Off.</u>	<u># Conf.</u>	<u>Max VOI</u>	<u>Act. VOI</u>	<u>% of Max</u>
.55	3	2	1	.39	.37	95
.60	2	3	3	.34	.31	91
.65	2	4	2	.29	.26	90
.70	2	2	2	.24	.24	100
.70	3	4	3	.24	.22	92
.70	3	4	3	.24	.19	79
.70	3	3	4	.24	.21	88
.70	4	4	4	.24	.19	79
.70	4	4	4	.24	.19	79
.80	2	4	2	.16	.14	88
.80	2	3	2	.16	.13	81
.80	3	4	3	.16	.13	81
.80	3	4	3	.16	.11	69
.80	2	3	3	.16	.13	81
.80	3	3	4	.16	.11	69
.85	4	3	4	.11	.08	73
.85	3	3	4	.11	.09	82
.95	3	3	4	.03	.02	67
1.00	3	3	4	.00	.00	--

Each row presents data for a single participant; they are ordered in increasing value of initial confidence.

As pointed out earlier, and as shown in the table, the higher the prior probability, the lower the maximum achievable VOI. As P reaches .85 and above, the VOI's become so small that the percent of maximum achieved becomes meaningless. If the data for participants with a P of .85 or higher are discarded, the maximum VOI's range from .16 to .39, the actual VOI's from .11 to .37, and the percent of maximum VOI achieved from 69 to 100. It should be

pointed out that, while maximum VOI and actual VOI depend heavily on the prior probability, the percent of maximum VOI achieved reduces the influence of prior probability and is a better reflection of the actual collection strategy employed (i.e., the extent to which diagnostic indicators were selected). Since our primary interest in this phase of the research is on collection strategy, we will use the percentage measure.

It is interesting to note that the only participant to achieve 100% of the maximum VOI employed a completely balanced collection strategy--that is, of the four indicators selected, two were in the hypothesized and two in the non-hypothesized sector; two were offensive and two defensive; two were confirming and two disconfirming. The actual VOI depends not only on the general collection strategy but on the specific indicators selected; this participant selected the most diagnostic indicators available, namely, second echelon movement in each sector, and preparation of minefields in each sector. His verbal report confirmed that he had in fact based his selections on the P values provided. His prior probability was a cautious .70, and he was able to achieve not only 100% of the maximum VOI, but a reasonably high absolute score of .24.

In an effort to determine how the achieved VOI related to general collection strategy, the Trial 1 data for 15 of the subjects (those whose prior probabilities were less than .85) were grouped in accordance with how many of their selected indicators were (1) in their hypothesized sector, (2) offensive, and (3) confirming, and a mean of the percent of maximum VOI achieved for each group was computed. Table 3-6 presents these data, and Figure 3-1 shows the trends in graphic form. It is clear from Table 3-6 and Figure 3-1 that as the number of indicators in each category goes up, the percent of maximum VOI achieved goes down. The high value of 94.9 obtained with one confirming indicator (i.e., three disconfirming) is an anomaly. It is based on a single case, a participant whose initial confidence level was a low .55 and who selected highly diagnostic indicators, thereby achieving a very high score. In general, we may interpret the tendency to select indicators in the hypothesized sector, and indicators that are confirming, as evidence of a bias in collection strategy. The data show that as bias increases, performance as measured by VOI goes down.

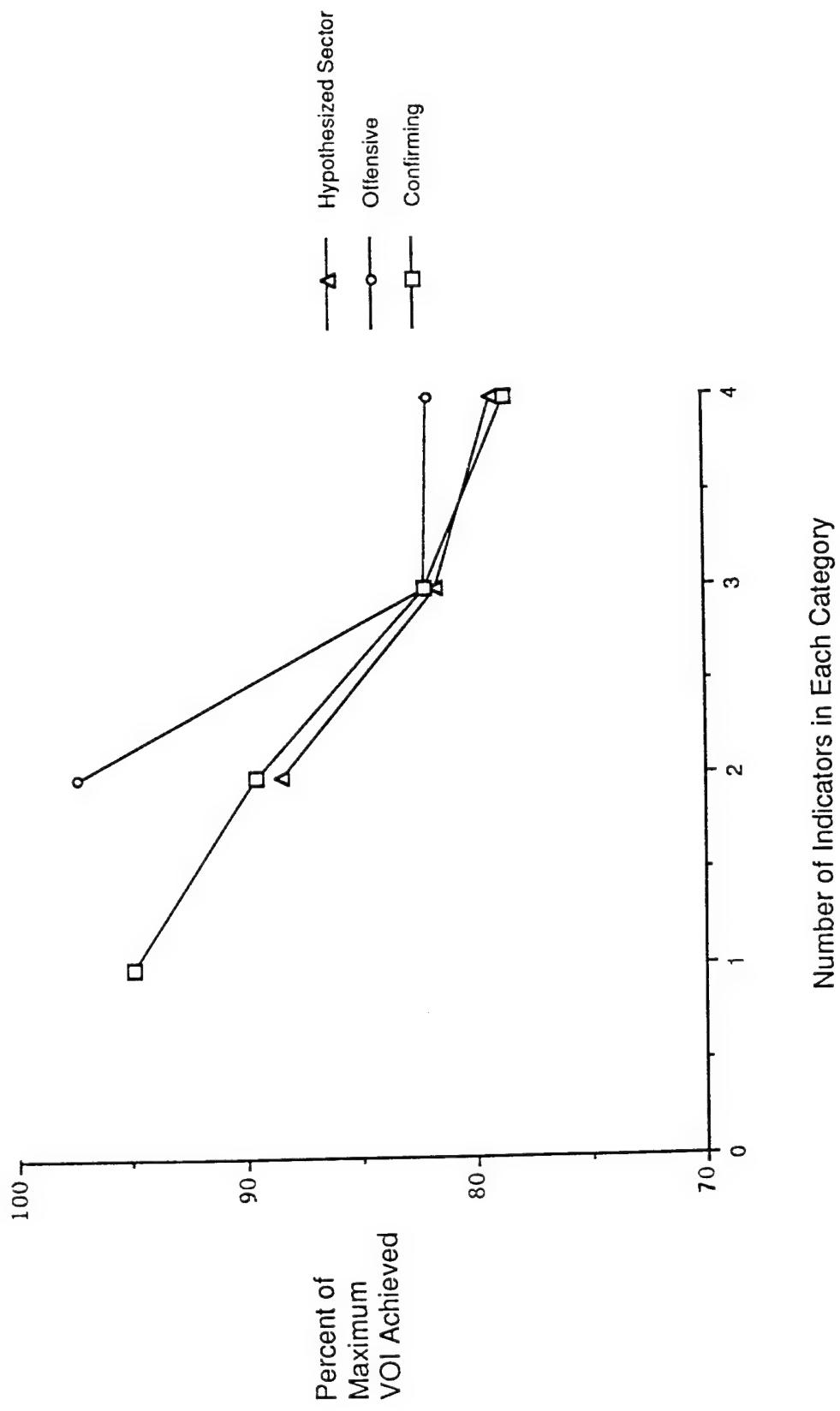


Figure 3-1. Relationship Between General Collection Strategy and VOI

Table 3-6: Average VOI (percent of maximum) as a Function of Collection Strategy

<u>Number of Indicators</u>	<u>In Hypoth Sector</u>	<u>Offensive</u>	<u>Confirming</u>
1	---	---	94.9*
2	88.5	97.4	89.6
3	81.7	82.2	82.2
4	79.2	82.0	78.6

*Based on 1 case, with low (.55) prior probability and selection of highly diagnostic indicators

3.5 Effects of Feedback on Collection Strategy and VOI

Table 3-7 presents the relevant Trial 2 collection strategy and VOI data for individual participants. For comparison with Table 3-5 (for Trial 1), the columns from left to right show:

confidence in the hypotheses after Trial 1 feedback
 (the prior probability);
 number of indicators chosen in the hypothetical sector;
 number of offensive indicators;
 number of confirming indicators;
 maximum VOI achievable for the assessed prior probability;
 actual VOI achieved;
 percent of maximum VOI achieved.

As in Table 3-5 the rows are in order of increasing confidence level, and again as P reaches .85 and above the VOI's become so small that the percent of maximum achieved becomes meaningless. Discarding data for participants with a P of .85 or higher, the maximum VOI's range from .13 to .28 (as compared with .16 to .39 in Trial 1), the actual VOI's from .06 to .23 (as compared with .11 to .37 in Trial 1), and the percent of maximum VOI achieved from 44 to 95 (as compared with 69 to 100). These values tend to be lower than those in Trial 1 because there were fewer highly diagnostic indicators remaining to choose from.

Table 3-7: Trial 2 Strategies and VOI for Individual Participants

Prior <u>Prob.</u>	# <u>Hyp.</u>	# <u>Off.</u>	# <u>Conf.</u>	Max <u>VOI</u>	Act. <u>VOI</u>	% of <u>Max</u>
.60	1	2	1	.28	.21	75
.65	1	3	2	.27	.22	81
.65	2	0	2	.25	.23	92
.65	2	2	4	.23	.17	74
.70	1	2	1	.21	.19	90
.70	3	4	3	.17	.13	76
.70	1	0	3	.22	.10	45
.70	1	2	1	.22	.21	95
.70	0	2	2	.22	.19	86
.75	2	1	3	.18	.08	44
.80	2	0	2	.13	.09	69
.80	2	1	1	.13	.12	92
.80	2	2	2	.13	.06	46
.80	2	2	2	.13	.12	92
.85	2	2	2	.10	.04	40
.85	1	1	2	.10	.05	50
.90	2	2	2	.06	.02	33
.95	2	2	2	.02	.01	50
1.00	1	3	0	.00	.00	--

Table 3-8 summarizes the differences between Trials 1 and 2 with regard to collection strategies and VOI. It shows the mean number of indicators chosen that were in the hypothesized sector, offensive, and confirming, as well as the mean VOI and % of maximum VOI achieved, in each trial. The data are consistent with those of Table 3-4, which showed a shift in strategy between Trials 1 and 2. On average, about 3 out of 4 indicators selected in Trial 1 were in the hypothesized sector, offensive, and confirming, while less than 2 out of 4 in Trial 2 show these characteristics. It is interesting to note that the individual who achieved 100% of maximum VOI using a completely balanced strategy and selecting the most diagnostic indicators, adopted a less balanced strategy in Trial 2 (3 in hypothesized sector, 4 offensive, and 3 confirming), and his VOI score dropped to 76%.

Table 3-8: Trial 1 vs. Trial 2 Strategies and VOI

	<u>Hyp</u>	<u>Off</u>	<u>Conf</u>	<u>VOI</u>	<u>% Max VOI</u>
Trial 1 Mean	2.84	3.32	3.00	.164 (.195*)	84.1*
Trial 2 Mean	1.58	1.74	1.95	.118 (.151*)	75.5*

*Excluding participants with prior P of .85 or greater

Table 3-8 shows that the average VOI as well as the average percent of maximum VOI were both lower in Trial 2 than in Trial 1, despite the general trend towards a more balanced strategy. As indicated earlier, this is because there were fewer highly diagnostic indicators among those remaining to choose from, and a high score requires not only a balanced strategy but selection of the most diagnostic indicators.

3.6 Effect of Experience on VOI

It might be hypothesized that participants with more experience would be expected to achieve higher VOI scores, by being more cautious in their confidence level as well as by adopting a more balanced collection strategy. Figure 3-2 shows the percent of maximum VOI achieved as a function of years in service. We have omitted the data for one participant, whose initial

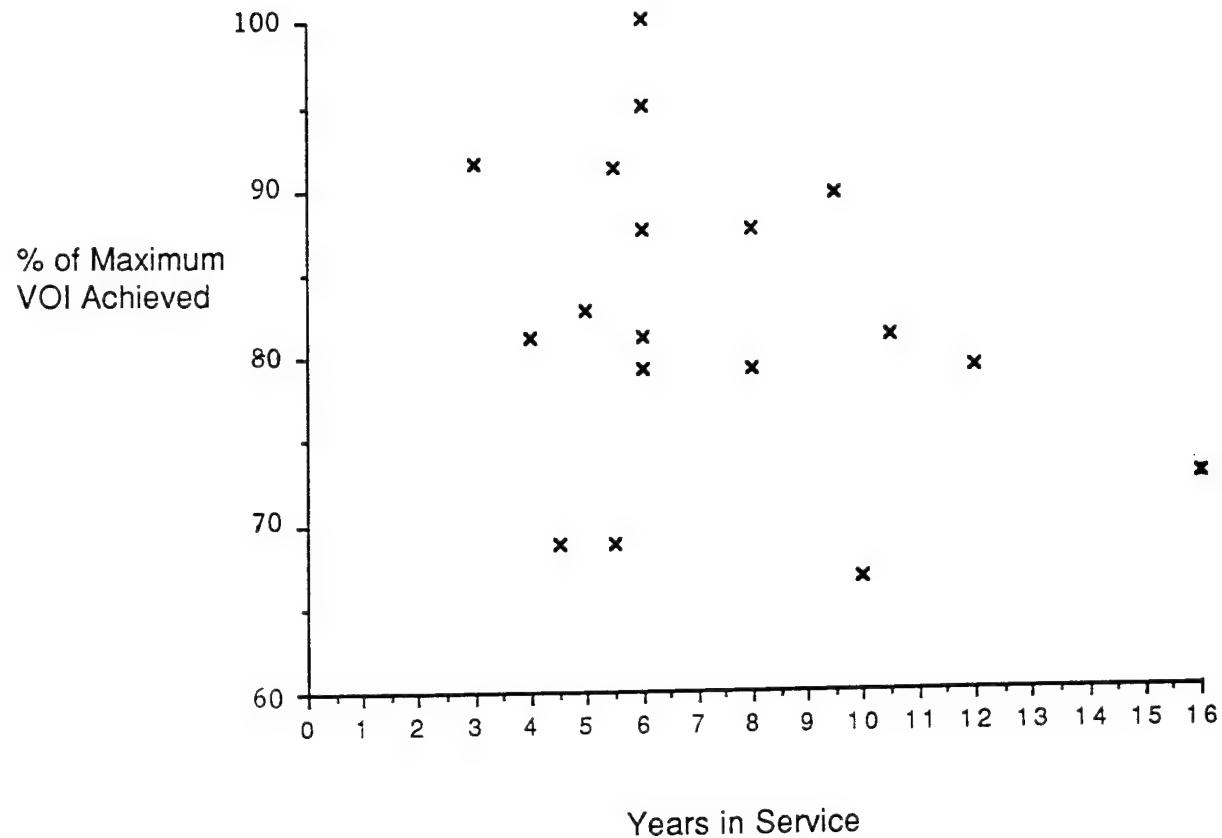


Figure 3-2. Percent of Maximum VOI Achieved as a Function of Years in Service

confidence level (prior probability) was 100%, and whose VOI score was therefore 0. A slightly negative relationship may be discerned, and the Pearson r is $-.30$, but the relationship is not statistically significant. Table 3-9 presents the data on years in service, years in Intelligence, and Percent of Maximum VOI achieved. For most of the participants, their years in Intelligence exactly equalled their years in service, but there were some exceptions. A Pearson r between percent of maximum VOI achieved and years in Intelligence likewise showed no relationship.

3.7 Effect of Feedback on Confidence Level

In Phases 1 and 2, when participants were passive receivers of ambiguous new items of information, their confidence levels tended to rise as the situation evolved, since confirming items were generally regarded as more important than disconfirming ones. In Phase 3 they selected a subset of four of the available indicators, and again received ambiguous information in the form of feedback, that is, two reports confirmed and two disconfirmed their initial hypothesis. For half the participants the confirming feedback was furnished for the indicators they had ranked 1 and 4, and the disconfirming for 2 and 3; for the other half this pattern was reversed. This design was an attempt to ensure that the feedback was balanced, i.e., that neither confirming nor disconfirming evidence was associated with indicators thought to be significantly more important.

As an independent check on this assumption, and as mentioned in Section 2-3, participants were asked to allocate 100 points of "value" among the four indicators selected in each of the two trials. This was to determine if the feedback was in fact balanced in terms of the perceived value of the indicators.

Table 3-10 shows how the weights were allocated by each participant in each trial. A t-test showed that the sum of the weights allocated to indicators 1 + 4 was not significantly different from those allocated to indicators 2 + 3, in either Trial 1 ($p = .096$) or Trial 2 ($p = .264$). Thus we may conclude that as far as the participants were concerned, the confirming and disconfirming feedback was applied to indicators that were about equal in perceived value.

Table 3-9: Relationship Between Experience and VOI Score

<u>Years in Service</u>	<u>Years in Intelligence</u>	<u>% of max VOI</u>
3	1.5	91.6
4	4	81.2
4.5	4.5	68.8
5	5	81.8
5.5	4.5	91.2
5.5	5.5	68.8
6	4	94.9
6	6	87.5
6	6	100.0
6	6	81.2
6	6	79.2
8	8	87.5
8	8	79.2
9.5	9.5	89.6
10	10	66.7
10.5	0.25	81.2
12	12	79.2
16	12	72.7

Correlation Coeff: % of max VOI, vs. Years in Service: -.304

% of max VOI, vs. Years in Intelligence: -.303

Table 3-10: Allocation of Weights to Indicators

<u>Participants</u>	Indicators:	<u>Trial 1</u>		<u>Trial 2</u>	
		<u>1 + 4</u>	<u>2 + 3</u>	<u>1 + 4</u>	<u>2 + 3</u>
1		50	50	50	50
2		50	50	50	50
3		50	50	50	50
4		50	50	50	50
5		55	45	50	50
6		45	55	50	50
7		50	50	50	50
8		50	50	50	50
9		55	45	50	50
10		50	50	50	50
11		55	45	60	40
12		40	60	40	60
13		47	53	45	55
14		50	50	45	55
15		55	45	45	45
16		50	50	50	50
17		82	18	97	3
18		50	50	40	60
19		60	40	60	40
Mean		52.3	47.7	52.2	47.8

Having assured ourselves that the feedback was balanced, we then examined the effect of the feedback information on confidence level. Table 3-11 presents the ranges and means of the confidence levels initially and after Trials 1 and 2.

Table 3-11: Effect of Feedback on Confidence Level

	<u>Initial Confidence</u>	<u>Confidence After Trial 1</u>	<u>Confidence After Trial 2</u>
Mean	76.05	76.57	76.57
Range	55-100	60-100	50-100

The table shows that, for the group, there was practically no change in confidence as the exercise progressed, the mean level remaining at about 76% throughout. This result differs substantially from that found in Phase 1, where confidence tended to rise as the exercise progressed. It also differs from that found in Phase 2, where confidence in general was lower than in Phase 1 but also rose slightly between the beginning and the end of the exercise.

For comparison purposes, Table 3-12 presents the trends in average confidence levels for each phase of the research.

Table 3-12: Trends in Average Confidence Levels, by Phase

	<u>Initial Confidence</u>	<u>1st Update</u>	<u>2nd Update</u>	<u>3rd Update</u>
Phase 1	77.3	79.6	82.2	80.0
Phase 2	67.0	62.3	64.7	71.9
Phase 3	76.1	76.6	76.6	*

*There were only two updating trials in Phase 3.

The difference between Phases 1 and 2 can be explained by the Phase 2 initial indoctrination on common decision biases and the graphic displays that

encouraged awareness of uncertainties. These interventions were not included in Phase 3, and the initial confidence level was almost as high as in Phase 1, as would be expected. However, in Phase 3 the participants were forced to assess the relative importance of the indicators in order to make their selections, and before receiving the feedback. Therefore they might have been just as influenced by contradictory evidence as by confirming evidence, and as a result, more inclined to interpret the feedback as truly ambiguous rather than as a justification for increased confidence. In short, they showed less bias in interpreting information that they themselves had identified in advance as important, than they did when passively receiving new information. It should be noted that this hypothesis was not tested directly by the research reported here, since the confidence judgments were obtained under somewhat different circumstances. For one thing, the items of information used as feedback in Phase 3 differed in some respects from those presented as intelligence reports in Phases 1 and 2. Secondly, participants worked in pairs during Phases 1 and 2, as compared with singly in Phase 3, which could possibly have affected their feelings of confidence. Finally, only two updating trials were given in Phase 3 as compared with three in Phases 1 and 2. It would be important to verify the finding by a direct comparison because of its important theoretical implications. The finding would also have practical applications for the processing of intelligence information in the field, as well as for training in collection management and interpretation of evidence.

It is interesting to note that the ranges of confidence levels shown in Table 3-11 extend to 100% at the upper end. This extreme level was given by one individual both initially and after Trial 1, but was dropped to 50% by that individual after Trial 2, whereas another individual, whose confidence had been 80% initially and 90% after Trial 1, expressed 100% confidence after Trial 2. These are clearly not typical responses. However, several of the participants expressed the feeling that officers at the G-2 and Division Commander level preferred that their staff show a high level of "decisiveness," which they interpreted as meaning confidence, in their judgments. If this is true, it suggests either a misunderstanding, calling for clarification of communications, or an ignorance of probabilistic judgments by high level personnel, calling for training appropriate personnel at higher levels.

4.0 CONCLUSIONS

4.1 Discussion of Findings

The major findings were as follows:

- a) Initially (i.e., in Trial 1) the strategy was to look for indicators in the hypothesized sector, that were offensive in nature, and that if found would confirm the hypothesis; there was also tendency to select indicators that would occur frequently regardless of the sector of enemy approach. Many indicators chosen were highly diagnostic, but the strategy ignored indicators that were even more diagnostic than some of those selected, and therefore resulted in lower Value of Information (VOI) scores than could otherwise have been achieved.
- b) In Trial 2 the tendency was reversed, although not enough to overcome the imbalance in the total frequencies; thus, the VOI scores in Trial 2 were lower than in Trial 1 because there were fewer highly diagnostic indicators remaining and those that did remain were not selected frequently enough.
- c) For the most part, the participants ignored the diagnosticity data, $P(E|N)$ and $P(E|S)$, which were provided to them; when diagnostic indicators were chosen, the selection was based on knowledge about enemy doctrine. Thus, whereas certain enemy defensive activities in one sector were characterized (by the experimenters) as highly diagnostic of an enemy attack in the other sector, such indicators tended to be ignored because defensive activities were not considered typical of enemy doctrine.
- d) Confidence levels did not rise over time, as they had in Phase 1 and to a smaller degree in Phase 2. We attribute this to the fact that in Phase 3 the participants selected and received feedback about the indicators they thought most important, rather than being the

passive recipients of information chosen by the experimenters. Our hypothesis is that under these conditions they would be less likely to discount evidence that disconfirms their early judgment, and since the feedback was balanced their confidence in their early judgment would be less likely to change. This explanation warrants a direct test, inasmuch as the Phase 3 conditions differed in some respects from those of Phases 1 and 2.

- e) The Value of Information (VOI) is a useful normative model for measuring performance in this type of research, although its limitations should be recognized.

These findings are consistent with those of Baron, Beattie and Hershey (1988), who found a tendency to overvalue questions that have a high probability of a positive result given the most likely hypothesis, which they refer to as a "congruence" bias. Other investigators (Fischhoff and Beyth-Marom, 1983; Tweney, Doherty, and Mynatt, 1982) who have obtained similar results, have termed it "confirmation bias." The confirmation behavior exhibited in this exercise is the result of a tendency to look for enemy offensive activity, and to ignore the possibility of finding such activity in the other sector, which would be equally diagnostic but would tend to disconfirm the initial hypothesis.

Although the tendency to focus on the expected sector of the enemy attack is not optimal in this exercise, the tendency to look for offensive rather than defensive activity can be justified by reference to known enemy doctrine which was not completely reflected in the diagnosticity data that we provided to the participants. For example, one of the very highly diagnostic indicators in the set provided was "preparation of minefields along the FLOT (forward line of troops)." This is a defensive action which would be extremely unlikely to occur (.05) in the sector in which an attack was planned, since mines would hinder the advance of the attacking troops. Yet it was selected in Trial 1 by less than one-third of the participants (6 out of 19), although one participant looked in both sectors. The reasons given for not selecting it were that (1) enemy doctrine stresses offensive rather than defensive activities, and (2) if defensive actions were taken, they would occur later

than offensive maneuvers. In fact, this indicator was selected more frequently in Trial 2, by 9 participants, four of whom looked in both sectors.

The important implication of this is that, in an evolving situation typical of battlefield operations, the diagnosticity of evidence is likely to change over time, partly as a result simply of the passage of time, but also very probably as a function of the changing situation. Any computer-based inference aid designed to facilitate situation assessment under dynamic conditions must take account of the changing diagnosticity of evidence. For some kinds of evidence the changes might be programmed to occur automatically with time, but it is likely that the more important changes would have to be based on events, possibly unanticipated, recognized by on-site experts, and the modifications introduced into the system manually. Since it would undoubtedly be unacceptably onerous to require military personnel to enter numerical probabilities into a system in real time, the system would have to be designed to calculate the probabilities on the basis of real-world events entered by on-site personnel. It would be naive to underestimate the complexity of the design problem implied by these considerations.

The training implications are perhaps easier to deal with. The findings suggest that the course in tactical intelligence should include material dealing with inferencing and decision making, the cognitive aspects of situation assessment, which to our knowledge are not now included. Perhaps the most important cognitive skills to be addressed are: formulating and testing hypotheses, recognizing uncertainties, gathering and interpreting information, and evaluating results. Diagnosticity of evidence, how new evidence should be incorporated with old, and the impact of new evidence on confidence levels, are vital concepts in the assessment of enemy intent. Intelligence analysts should be made aware of Bayesian and other approaches to the management of knowledge and uncertainty, not necessarily to turn them into Bayesian statisticians but to imbue them with an understanding of the relative importance of various factors that should affect confidence level. They should be taught to recognize the heuristics or short-cuts that are often used in the reasoning process, how they can be useful and when they may lead to so-called "biases" or errors in judgment.

These concepts should be presented in the context of scenarios that are currently used as practical exercises, and these exercises should be modified to represent evolving battlefield situations, so that students can learn how to assimilate new information with old and how to assess the diagnosticity of evidence as the situation changes. In particular, exercises should be designed to illustrate how the field of attention narrows during stress (time stress can be used for exercise purposes), how base rate information may be ignored or inappropriate base rates used, and the various ways in which the confirmation bias may evidence itself during an evolving situation, in situation assessment and in collection management for testing hypotheses.

4.2 Recommendations

The recommendations that follow are based on the findings of all three phases of this project.

4.2.1 Training implications. The Phase 1 and 2 findings showed that trained intelligence analysts are less likely to ignore or undervalue evidence that contradicts their early hypothesis if they understand that the tendency to do so is a commonly found cognitive bias that can undermine their judgment. The most serious potential consequence of this bias is an increasing confidence in an early judgment that may be in error; such a bias, one of the contributing factors in the USS Vincennes downing of an Iranian airbus in July 1988, was referred to in the official Navy report on the incident (U.S. Navy, 1988) as "scenario fulfillment." It would, of course, be undesirable for a training program to produce intelligence analysts who cannot make up their minds. Rather, the objective would be to ensure that they give appropriate attention to disconfirming as well as to confirming evidence when new information is received.

Closely related to this issue is that of confidence in judgment. In addition to the unwarranted increase in confidence due to undervaluing contradictory evidence, there appears to be tendency, overtly expressed by several of the participants, to believe that senior officers (e.g., Division Commanders, G-2's) are uncomfortable with the indecision implied by low confidence levels. If this perception is accurate, it suggests the need for training of senior

officers in the potential degradation in their staff's judgmental performance that this attitude may cause. If the perception is incorrect, it suggests the need for training of senior officers in the importance of communicating a more appropriate point of view to their staff. What is most likely is that it accurately reflects the attitude of some but not all senior officers, suggesting that they all should be exposed to both approaches, i.e., the need to "keep an open mind," and the importance of communicating this to their intelligence staff.

Finally, as suggested by the findings in all phases of the work, there is a general need to incorporate into the Officer Advanced Course at USAICS some instructional and exercise material dealing directly with the diagnosticity of intelligence indicators and techniques for incorporating new evidence, or the lack thereof (negative information), into inferential judgments in an evolving and uncertain situation. Although the use of rigorous Bayesian updating procedures may be onerous and inappropriate in a combat situation, student exposure to these concepts, embedded in realistic battlefield exercises, should provide a degree of understanding that can be effectively applied in a qualitative way during combat. Furthermore, if battlefield aid developments succeed in producing a situation assessment aid based upon Bayesian formulas, it will be essential for users of such an aid to understand the basis of its recommendations in order to ensure its credibility as well as to permit them to work with it (i.e., assess its conclusions, modify its inputs, modify or ignore its outputs).

Therefore, we recommend that:

- a) Students in the Officer Advanced Course at USAICS be given training in typical cognitive biases that can undermine judgment, especially the various forms of confirmation bias (seeking confirming indicators, ignoring or undervaluing contradictory evidence, seeking highly likely rather than diagnostic indicators), and in the concepts and techniques of Bayesian inference, including the criteria of diagnosticity, the updating of judgments, and the factors that may cause the diagnosticity of indicators to change as a battlefield situation evolves. Examples illustrating these

concepts should be presented in the context of exercises already being conducted during the course work; these exercises may have to be modified slightly to accommodate the notion of an evolving battlefield situation with new evidence being received after early hypotheses have been formulated, but these modifications are not likely to be extensive. A brief course for instructors, and supplementary reference and syllabus material, should be prepared to ensure that the existing training practices can be maintained.

- b) Consideration should be given to incorporating into the training of senior officers a brief exposure to the concepts described above, but with emphasis on the potential consequences of demanding or expecting invariably high confidence levels in their staff's judgments, and in the importance of communicating to their staffs the guidance to "keep an open mind." This recommendation cannot be made as strongly as the one above, since the justification for it is more speculative and possibly could be further investigated before it is put into effect.

4.2.2 Decision aiding implications. The implications for decision aiding derive mainly from Phases 1 and 2, and relate primarily to the use of graphic techniques to promote an awareness of the uncertainties inherent in the battlefield situation and to provide help in coping with these uncertainties. The Phase 3 results support the potential usefulness of an inference aid but point up the difficulty of designing such an aid in the face of changing diagnosticities of indicators as the situation evolves.

Recommendations for graphic decision aids are as follows:

- a) Graphic Enemy Order of Battle (EOB) Display

A major source of uncertainty in situation assessment is the location of enemy units that are known (or assumed) to exist, as components in the EOB, but have not yet been located. Currently, EOB is provided in tabular form in a Workbook, and located enemy units are plotted on a map overlay. Intelligence analysts were

found in this project to work primarily at the map display, as a result of which they often made inferences based on the plotted enemy units and ignored the unlocated units listed in the Workbook. In Phase 2, a graphic EOB display was provided adjacent to the map, showing organization and unit strength, and making clear which units were still unlocated. This display made the uncertainties obvious, and was found very useful by all participants. A computerized version, keyed to a computer-based map display, is recommended for development.

b) EOB-by-Equipment Display

The EOB Workbook is traditionally organized by unit. Thus, if an item of equipment is located on the battlefield, the analyst must scan through the Workbook to identify units that might own the equipment and determine how many such items each unit owns (base rates that would assist in inferring ownership). This is an onerous procedure that was rarely followed. In Phase 2 participants were provided with a Workbook organized by equipment, so that they could turn to the equipment that was located and receive a display of all units that own such an item and how many they own (i.e., base rates), and thus infer quickly the probability that the equipment may belong to any particular unit. Because of inadequate design, this aid was not used very often and therefore did not receive an appropriate test. The concept, however, appears worthy of preliminary development and evaluation as a computerized data retrieval system.

c) Computerized Dynamic Event Templating

Event templating is a technique designed to encourage analysts to keep alternative hypotheses about enemy intentions under active consideration as the situation develops. It involves plotting on an overlay the location of named areas of interest (NAI's), such as road junctions and river crossings, and critical events that would be expected to occur if the enemy chose any of several possible

avenues of approach (AOA's). A related display, the events analysis matrix, consists of tabularly formatted distances and estimated time differences between successive NAI's (time and distance information can also be plotted on the event template). These data are based on rules of thumb about rate of movement of various types of unit under different terrain and weather conditions. Although event templating is one of the few, if not the only, procedure that forces the situation assessor explicitly to consider alternative enemy actions, it was described as difficult to accomplish, unfamiliar, and not very helpful. Even when pre-drawn event templates were made available (in Phase 2), they were rarely used.

A computer-driven dynamic map display that permits the user to project enemy activity under alternative assumptions about AOA, would seem to be a potentially valuable component of a situation development aid. A possibly useful feature would be the capability to run the projection in fast time, to facilitate anticipation and perception of movement patterns. The concept of a computerized dynamic event template appears worthy of further exploration.

d) Prompting to Reduce Bias

In view of the strong tendency to regard confirming information as more important than disconfirming, especially if indicators are not identified as important in advance, one other aiding technique seems worth investigating. An AI system to aid in situation assessment would presumably incorporate in its knowledge base a set of indicators of enemy intentions (e.g., of offensive and defensive activity, of various avenues of approach, etc.). In an evolving situation, analysts would tend to look for indicators confirming their early judgments, and either ignore or perhaps misinterpret contrary evidence. A technique that might help counteract this tendency would be to have the analyst enter his early judgment into the system, and have the system highlight new information that tends to contradict that judgment, to call it to the analyst's attention. Even if the strength of the evidence cannot be quantified in advance

but must be judged in context by the analyst, the attention-getting characteristic of this technique may be an effective countermeasure to the confirmation bias.

4.2.3 Research implications. Two areas of further research are suggested:

- a) The finding that confidence level tended not to rise in Phase 3 when the type of evidence thought to be important was selected by the participants, deserves further investigation. This result contrasts with Phases 1 and 2, when participants had no influence over the evidence furnished to them by the experimenters. Since the evidence in all phases was balanced with regard to the two hypotheses, north and south, the findings suggest that intelligence analysts would pay more attention to contradictory evidence if they had identified that type of evidence as important in advance. In effect, their active selection of an indicator seems to commit them to the value of the information it provides, even if it contradicts their hypothesis, while their passive receipt of information leaves them more willing to discount it. However, the conditions in Phase 3 differed in several respects from those in Phases 1 and 2, and were not designed for a direct test of this hypothesis. It would be of theoretical interest to test this directly. If the findings are confirmed, there might also be some practical implications. Currently the tasks of collection management (selection of indicators) and situation development (assessment of evidence) are conducted by separate staff sections. A technique for reducing the confirmation bias might be to adopt a procedure in which the two functions are combined, or at least conducted coordinately. In fact, any procedure which encourages a situation developer to assess the diagnosticity of various indicators in advance of receiving intelligence reports would accomplish this result. The training recommendations discussed above (Section 4.2.1) would facilitate such procedures.
- b) The recommendation made earlier (Section 4.2.2.d) regarding an interactive system that prompts by highlighting contradictory

evidence, deserves further investigation and possible extension. In general, AI and expert systems have been dependent upon knowledge bases and control structures that have been formulated with regard to the problem domain of interest. Their advice to the user has taken the form of recommendations relevant to the solution of the problem. Except for some intelligent tutoring systems, which develop a model of the student's level of knowledge based upon the student's responses, AI systems tend not to incorporate a knowledge base about human cognition, and certainly are not designed to prompt users away from common cognitive biases. An exception is found in work by Cohen, Laskey and Tolcott (1986) and Cohen, Tolcott and McIntyre (1987), in which decision aids are designed not only to provide an "expert" solution to a problem if desired, but also to let the user solve the problem in his own fashion (personalization) while providing prompts of various kinds to steer the user away from sub-optimal procedures and biased solutions (prescription). The recommendation referred to above (i.e., highlighting contradictory evidence) is a step toward the design of a personalized and prescriptive aid for intelligence analysis.

In order to provide a firmer basis for exploiting this concept, basic research should be undertaken to increase our understanding of the task conditions under which decision performance occurs that might be improved by a combined personalized and prescriptive approach. It would be important to conduct this research in naturalistic settings, using reasonably trained personnel performing tasks for which they have been trained. The heuristics and rules of thumb learned during training or on the job can be very useful in dealing with high workload or time stress situations; the research issue is to identify the conditions under which these procedures could lead to significant errors because of cognitive biases that are now fairly well understood and avoidable. Research of this type would provide (1) a richer theoretical understanding of the decision dynamics of trained personnel, especially the performance benefits and costs of heuristic reasoning employed by them, and (2) a basis for a more applied program to develop and evaluate decision aids.

that allow use of preferred procedures while guarding against the errors that such procedures may result in.

REFERENCES

Cohen, M.S., Laskey, K.B., and Tolcott, M.A. *A personalized and prescriptive decision aid for choice from a database of options*. (Technical Report 86-1). Falls Church, VA: Decision Science Consortium, Inc., March 1986.

Cohen, M.S., Tolcott, M.A., and McIntyre, J.R. *Display techniques for pilot interactions with intelligent avionics: A cognitive approach* (Technical Report 87-6). Falls Church, VA: Decision Science Consortium, Inc., April 1987.

Baron, J., Beattie, J., and Hershey, J.C. *Heuristics and biases in diagnostic reasoning. II. Congruence, information, and certainty*. *Organizational Behavior and Human Decision Processes*, 42(1), 88-110, 1988.

Einhorn, H.J. Learning from experience and suboptimal rules in decision making. In T.S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1980.

Einhorn, H.J., and Hogarth, R.M. Confidence in judgment: Persistence in the illusion of validity. *Psychological Review*, 85, 395-416, 1978.

Fischhoff, B. and Beyth-Marom, R. Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, 90, 239-260, 1983.

Mynatt, C.R., Doherty, M.E., and Tweney, R.D. Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29, 85-95, 1977.

Tolcott, M.A., Marvin, F.F., Lehner, P.E. *Effects of early decisions on later judgments in an evolving situation* (Technical Report 87-10). Falls Church, VA: Decision Science Consortium, Inc., July 1987.

Tolcott, M.A., Marvin, F.F., and Lehner, P.E. Expert decision making in evolving situations. *IEEE Transactions on Systems, Man and Cybernetics*, 19, 606-615, 1989.

Tolcott, M.A. and Marvin, F.F. Reducing the confirmation bias in an evolving situation (Technical Report 88-11). Reston, VA: Decision Science Consortium, Inc., August 1988.

Tweney, R.D., Doherty, M.E., and Mynatt, C.R. Rationality and disconfirmation: Further evidence. *Social Studies of Science*, 12, 435-441, 1982.

U.S. Navy; Commander in Chief, U.S. Central Command. *First Endorsement on Rear Admiral Fogarty's letter of 28 July 1988* (Subj: Formal investigation into the circumstances surrounding the downing of Iran Air Flight 655 on 3 July 1988).

Wason, P.C. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 1960, 12, 129-140.

Wason, P.C. 'On the failure to eliminate hypotheses...' - a second look. In P.C. Wason and P.N. Johnson-Laird (Eds.), *Thinking and reasoning*, Middlesex, England: Penguin Books, Ltd., 1968, 165-174.

APPENDIX A

A.1 The Value of Information Model

In order to evaluate performance during the experiments, we required a normative model with which we could compare actual behavior. We used a Value of Information (VOI) model similar to the normative model described in Baron, Beattie, and Hershey (1988). In the simplest form we are evaluating a single question (is the attack more likely to be in the North or in the South) and we can collect evidence (e.g., has there been movement of artillery forward in the North?). Once we decide where the attack is more likely, we would typically defend more strongly in that sector. As a simplifying assumption, we only care if we are right or wrong, and we can associate a utility value of 1 with being correct and 0 with being wrong. Thus, if we choose to defend in the North and the attack comes in the South, this would be 0 on our utility scale. If we defend North and the attack is in the North, this would be a 1. Presumably we would always choose the course of action that corresponded to our assessment of the most likely sector for attack. We can represent this in decision tree form as shown in Figure A-1.

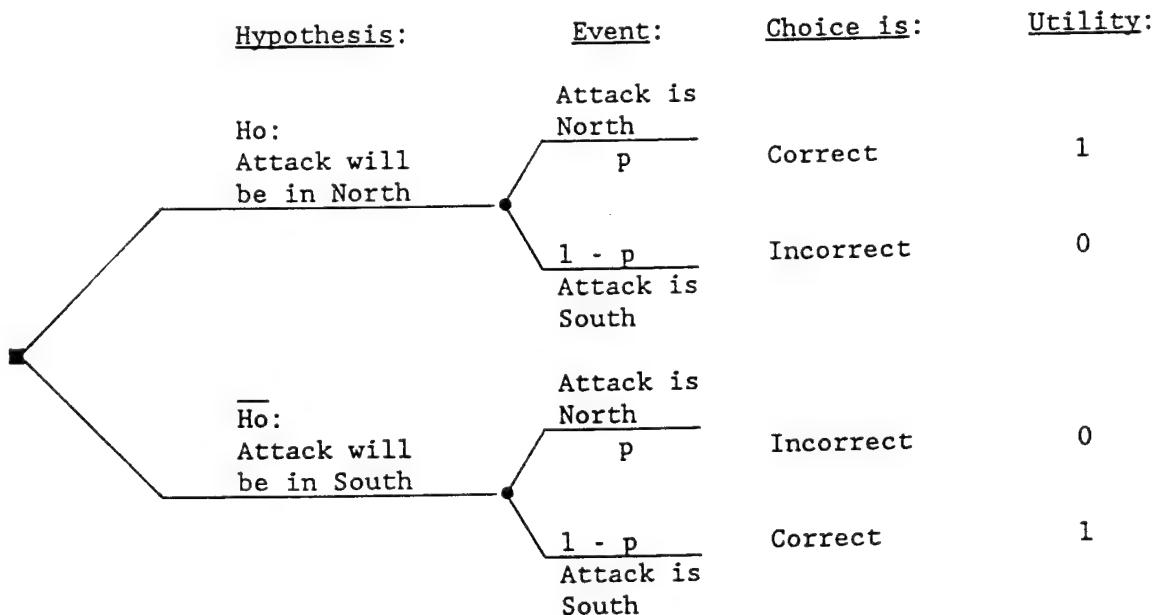


Figure A-1: Decision Tree Form of Value Of Information Model

Before evidence is gathered, p represents the prior probability that the attack will be in the North. With the utility scale described above, the expected utility of an alternative is equivalent to the probability of the most likely hypothesis. After evidence is gathered, we again calculate the expected utility of each hypothesis which again is the probability of selecting the correct choice. The Value of Information is, therefore, the difference between expected utility (or probability of the most likely hypothesis) with the evidence and without it. This is best illustrated by a sample example.

Example

H_0 : Attack will be in North (therefore, defend North) - N

H_1 : Attack will be in South (therefore, defend South) - S

$P(N)$ - Prior probability of N - p

$P(S)$ - Prior probability of S - 1 - p

Evidence - e

$P(e|N)$ - Probability of evidence e if attack is North - q

$P(e|S)$ - Probability of evidence e if attack is South - r

$P(\bar{e}|N)$ - Probability of not evidence e, if attack is North - 1-q

$P(\bar{e}|S)$ - Probability of not evidence e, if attack is South - 1-r

Initially, we can represent the prior probabilities and the probabilities of evidence conditioned on location of attack as shown in Figure A-2.

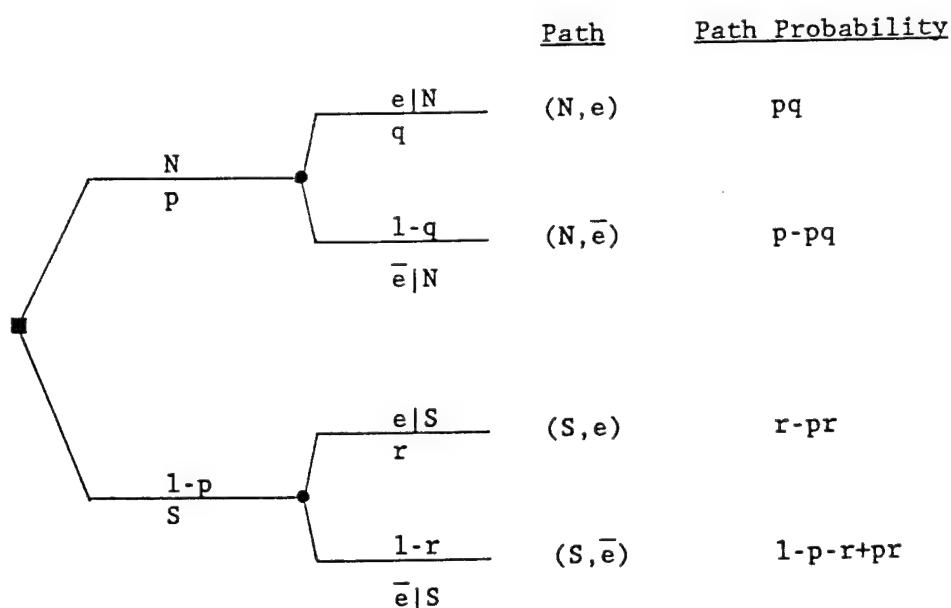


Figure A-2: Probabilities of Evidence Conditioned on Location of Attack

Since we will always act in response to the higher probability, we will be wrong at a frequency equal to the lower probability. If we assume $p > 1-p$, the probability of being wrong is $1-p$. The purpose of gathering evidence is to reduce this probability, and a normative approach would be to select evidence that reduces it the most. Using our utility scale, the value of information (VOI) of the evidence becomes the amount the probability of error is reduced.

Thus, to determine the VOI of evidence e , we need to calculate posterior probabilities of N and S conditioned on e . We do this by "flipping" the tree, assigning the relevant path probabilities (from Figure A-2) and calculating the probability of North given e , $p(N|e)$, North given not e , $p(N|\bar{e})$, the probability of South given e , $p(S|e)$, and South given not \bar{e} , $p(S|\bar{e})$; we then divide the path probability by the corresponding probability of evidence to calculate posterior probabilities. Figure A-3 shows this procedure in tree form.

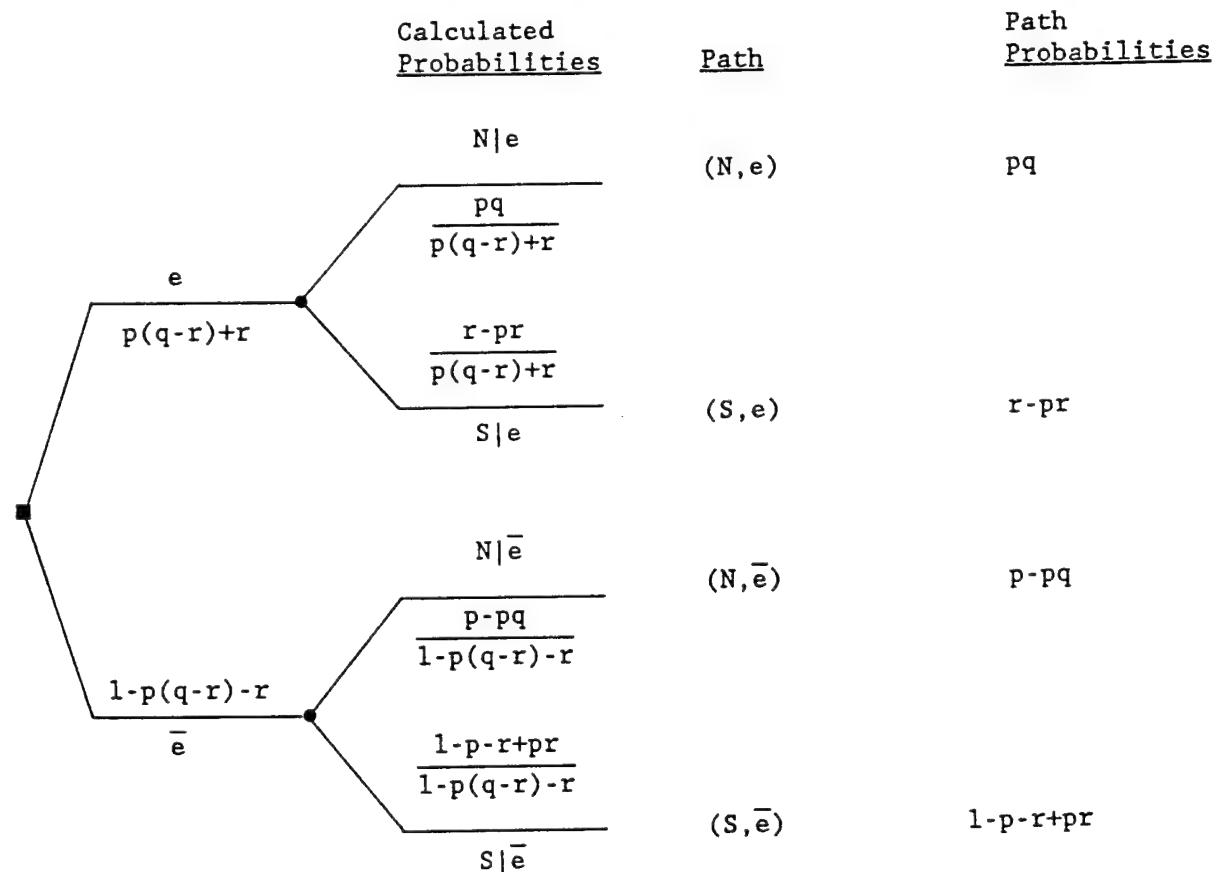


Figure A-3: Posterior Probabilities of Hypotheses Conditioned on Evidence

The probability for e shown above is obtained by summing probabilities along all paths where e occurs; that is, e occurs on path (N, e) with path probability of pq , and on path (S, e) with path probability $r-pr$. Summing probabilities for e we get $pq+r-pr = p(q-r) + r$. Similarly, probability for \bar{e} = $p-pq + 1-p-r+pr = 1-p(q-r)-r$.

To simplify, assume $p = .6$, $q = .6$, and $r = .2$. We then get probabilities for e and \bar{e} that can then be modeled as in Figure A-4.

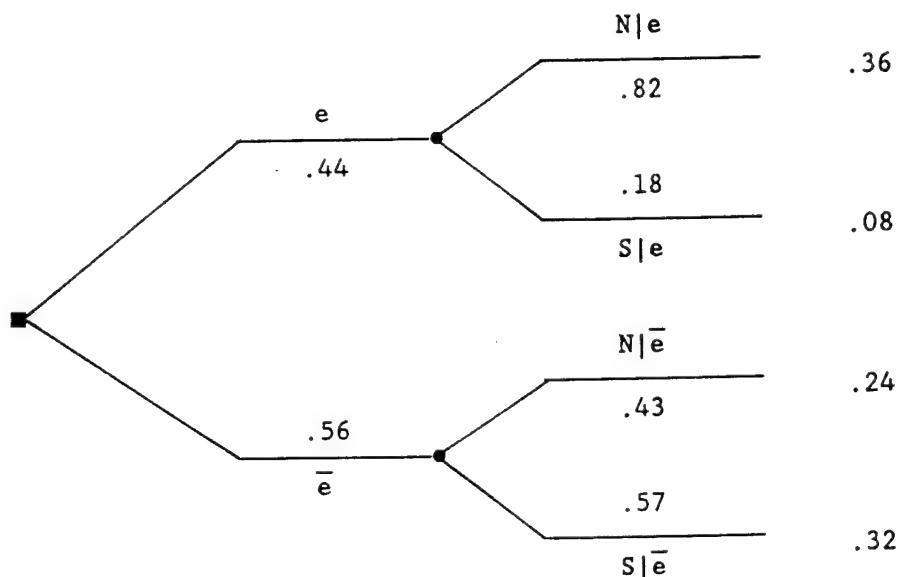


Figure A-4: Model for Probabilities of e and \bar{e}

The prior probability for North was .6 and for South was .4. If action were taken based upon the higher probability, with no other information, we would defend North with probability of error of .4. If e occurs, the more likely event is North since $.82 > .18$; by defending North, the probability of error is .18 and this will occur 44% of the time. If \bar{e} occurs, South is more likely since $.57 > .43$. The decision will be to defend South and the probability of error is .43, and this will occur 56% of the time. Thus, the expected error is:

$$(44\%) \times .18 + (56\%) \times .43 = .32$$

Since the original error probability was .40, and evidence e reduces it to .32, the VOI for e is $.40 - .32 = .08$. It is important to note that in this formulation the VOI depends on the prior probabilities; the higher the prior for the preferred hypothesis, the less value any new evidence can have. Thus, if the expressed prior for the preferred hypothesis is, say, .90, then the error probability is .10, and very little reduction in error is possible, regardless of the evidence.

Extending this to our experimental case in which we need the VOI for a "package" of four pieces of evidence at the same time, we can illustrate how these calculations would appear as shown below:

Assume the following for evidence a, b, c, and d:

$p(N)$	-	.6			
$p(S)$	-	.4			
$p(a N)$	-	.50	$p(a S)$	-	.90
$p(b N)$	-	.40	$p(b S)$	-	.30
$p(c N)$	-	.60	$p(c S)$	-	.20
$p(d N)$	-	.05	$p(d S)$	-	.65

The probability trees are as shown in Figures A-5 and A-6. [Note: Notationally, an "n" before a letter in the tree means "not"; thus "na" means not a].

Since each piece of the four pieces of evidence can be present or absent, there are 2^4 or 16 combinations of evidence as shown in the tree. The column of numbers to the left of the letters representing the pieces of evidence are the conditional probabilities of the package of evidence given North or South. The pieces of evidence are assumed to be independent; therefore, the probability is just the product of the individual probabilities. For example, for the branch "a, b, c, nd" given North, we calculate the probability of the evidence package given North as:

$$p(a|N) \times p(b|N) \times p(c|N) \times p(nd|N) = \\ .50 \times .40 \times .60 \times .95 = .114$$

	Prior Probability of Sector	Probability of Evidence/ Sector	Path Probability
1	0.018 a,b,c,d	0.204 NORTH 0.796 SOUTH	0.0036 0.0160
2	0.076 a,b,c,nd	0.900 NORTH 0.100 SOUTH	0.0684 0.0075
3	0.059 a,b,nc,d	0.041 NORTH 0.959 SOUTH	0.0024 0.0561
4	0.038 a,nb,c,d	0.142 NORTH 0.858 SOUTH	0.0054 0.0327
5	0.005 na,b,c,d	0.698 NORTH 0.302 SOUTH	0.0036 0.0015
6	0.076 a,b,nc,nd	0.601 NORTH 0.399 SOUTH	0.0456 0.0302
7	0.120 a,nb,c,nd	0.853 NORTH 0.147 SOUTH	0.1026 0.0176
8	0.069 na,b,c,nd	0.988 NORTH 0.012 SOUTH	0.0684 0.0008
9	0.135 a,nb,nc,d	0.027 NORTH 0.973 SOUTH	0.0036 0.1310
10	0.009 na,b,nc,d	0.278 NORTH 0.722 SOUTH	0.0024 0.0062
11	0.009 na,nb,c,d	0.597 NORTH 0.403 SOUTH	0.0054 0.0036
12	0.139 a,nb,nc,nd	0.492 NORTH 0.508 SOUTH	0.0684 0.0705
13	0.049 na,b,nc,nd	0.931 NORTH 0.069 SOUTH	0.0456 0.0033
14	0.105 na,nb,c,nd	0.981 NORTH 0.019 SOUTH	0.1026 0.0019
15	0.018 na,nb,nc,d	0.198 NORTH 0.802 SOUTH	0.0036 0.0145
16	0.076 na,nb,nc,nd	0.897 NORTH 0.103 SOUTH	0.0684 0.0078

Figure A-5: Probability Tree for $p(\text{Evidence}|\text{Sector})$

Probability of Evidence	Posterior Probability of Sector/ Evidence	Path Probability
1	0.006 a,b,c,d	0.004
2	0.114 a,b,c,nd	0.068
3	0.004 a,b,nc,d	0.002
4	0.009 a,nb,c,d	0.005
5	0.006 na,b,c,d	0.004
6	0.076 a,b,nc,nd	0.046
7	0.171 a,nb,c,nd	0.103
8	0.114 na,b,c,nd	0.068
NORTH		
9	0.006 a,nb,nc,d	0.004
10	0.004 na,b,nc,d	0.002
11	0.009 na,nb,c,d	0.005
12	0.114 a,nb,nc,nd	0.068
13	0.076 na,b,nc,nd	0.046
14	0.171 na,nb,c,nd	0.103
15	0.006 na,nb,nc,d	0.004
16	0.114 na,nb,nc,nd	0.068

		1.000
1	0.035 a,b,c,d	0.014
2	0.019 a,b,c,nd	0.008
3	0.140 a,b,nc,d	0.056
4	0.052 a,nb,c,d	0.033
5	0.004 na,b,c,d	0.002
6	0.076 a,b,nc,nd	0.030
7	0.044 a,nb,c,nd	0.018
8	0.002 na,b,c,nd	0.001
SOUTH		
9	0.328 a,nb,nc,d	0.131
10	0.016 na,b,nc,d	0.006
11	0.009 na,nb,c,d	0.004
12	0.176 a,nb,nc,nd	0.071
13	0.008 na,b,nc,nd	0.003
14	0.005 na,nb,c,nd	0.002
15	0.036 na,nb,nc,d	0.015
16	0.020 na,nb,nc,nd	0.008

Figure A-6: "Flipped" Probability Tree for $p(\text{Sector}|\text{Evidence})$

The column of numbers to the right is the path probability along each branch and is calculated as the product of the prior probability of the appropriate sector of attack and the probability of evidence given the sector. For example, for the path [NORTH, (a, b, nc, d)] we get:

$$p(N) \times p(a, b, nc, d) =$$

$$.6 \times .004 = .0024$$

Note that the chart reflects values rounded to the third decimal place.

In flipping the tree in Figure A-6, the rightmost column of numbers are the path probabilities. These are found by selecting the probability associated with each path in Figure A-5 and associating it with the path that has the same events. The column labeled Probability of Evidence in Figure A-6 is obtained by adding the path probabilities of all places where the package of evidence occurs. For example, evidence (a, b, c, d) occurs in Figure A-5 in the first path for North with probability .004 and in the first path for South with probability .014. Therefore, (a, b, c, d) occurs with probability .004 + .014 = .018 as shown in Figure A-6.

The column labeled Posterior Probability of Sector/Evidence is obtained by dividing the path probability by the probability of evidence.

For example, for North conditional on (a, b, c, d), we get :

$$.0036 + .018 = .20 \quad \text{[Note: the figure shows .204, the difference is due to round-off]}$$

What we are really after is the posterior probability for each sector based upon the evidence. The expected value of the posterior probability for North is the sum of all path probabilities where North occurs in Figure A-6, and would be .84; similarly the posterior for South would be .16. Since the probability for North is higher, we would defend North with an error probability of .16. The original (no evidence) error probability was .40; therefore, the VOI for the "package" of evidence a, b, c, and d would be .24.

In our experiment there are 16 pieces of evidence from which four can be selected (in Trial 1), thus there are $\frac{16!}{4! 12!} = 1,820$ combinations of four

items. We can evaluate all combinations as shown above, and the normative model thus becomes one of selecting the "package" with the highest VOI. We can similarly define other strategies including those that would be representative of the confirmation bias, calculate the VOI for each strategy, and make comparisons among generic behavior patterns.

A.2 Limitations of the VOI Model

As an absolute value the VOI is very sensitive to the prior probability (in this case, the expressed confidence in the initial hypothesis); the higher the prior probability, the lower the maximum VOI achievable). When the primary interest is on the information collection strategy, the influence of the prior probability can be minimized by using the percent of maximum VOI achieved, as the performance measure. However, even this measure loses its meaning with priors of about .85 and above, because the maximum achievable is so low that slight changes in VOI achieved have disproportionate effects on the percentage. Another disadvantage is that when several indicators are chosen simultaneously (in this case, four), computation of VOI (both maximum and actual) is a tedious process, although certainly feasible with existing computer programs. Finally, the actual diagnosticity of an indicator on the battlefield is likely to change with time, either as a function of time itself or of the changing situation. Thus the VOI of any single indicator or group of indicators should be recomputed as the battlefield situation evolves, in order to reflect such change.